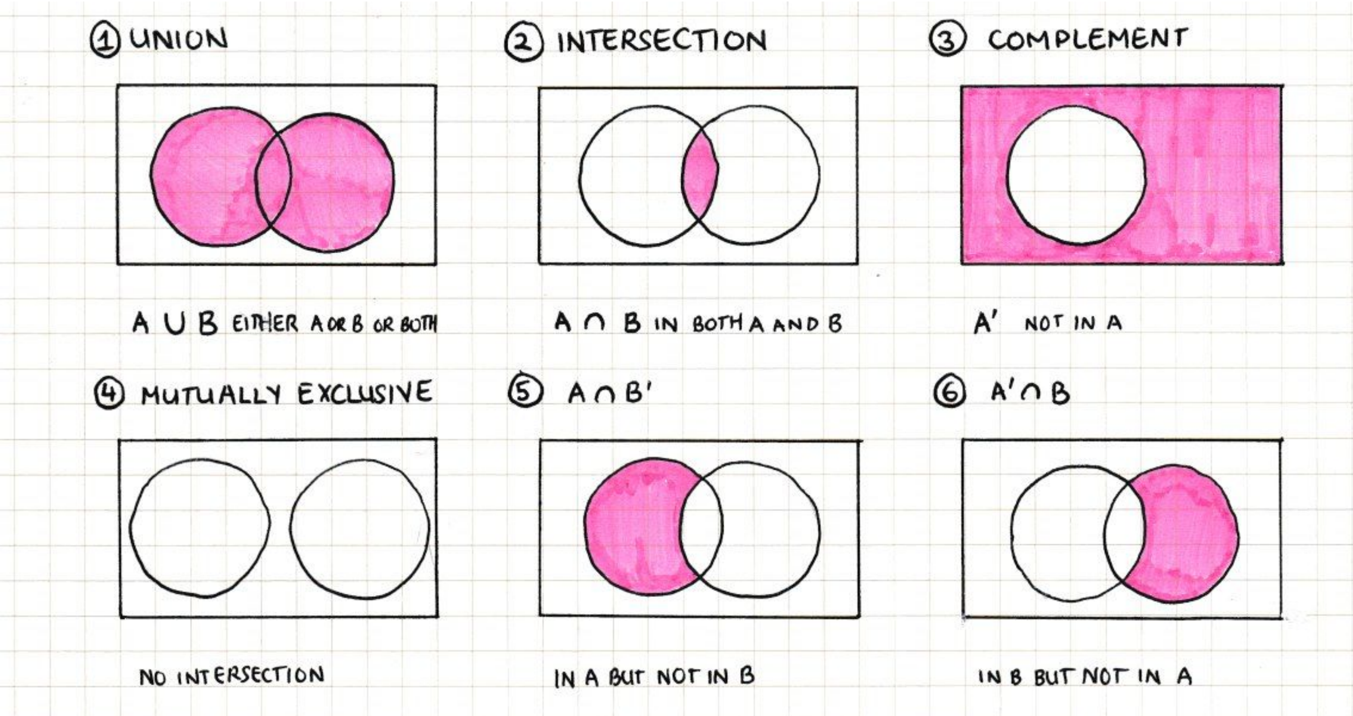
Conditional Probability
&
Bayesian Inference

Uohna Thiessen, PhD

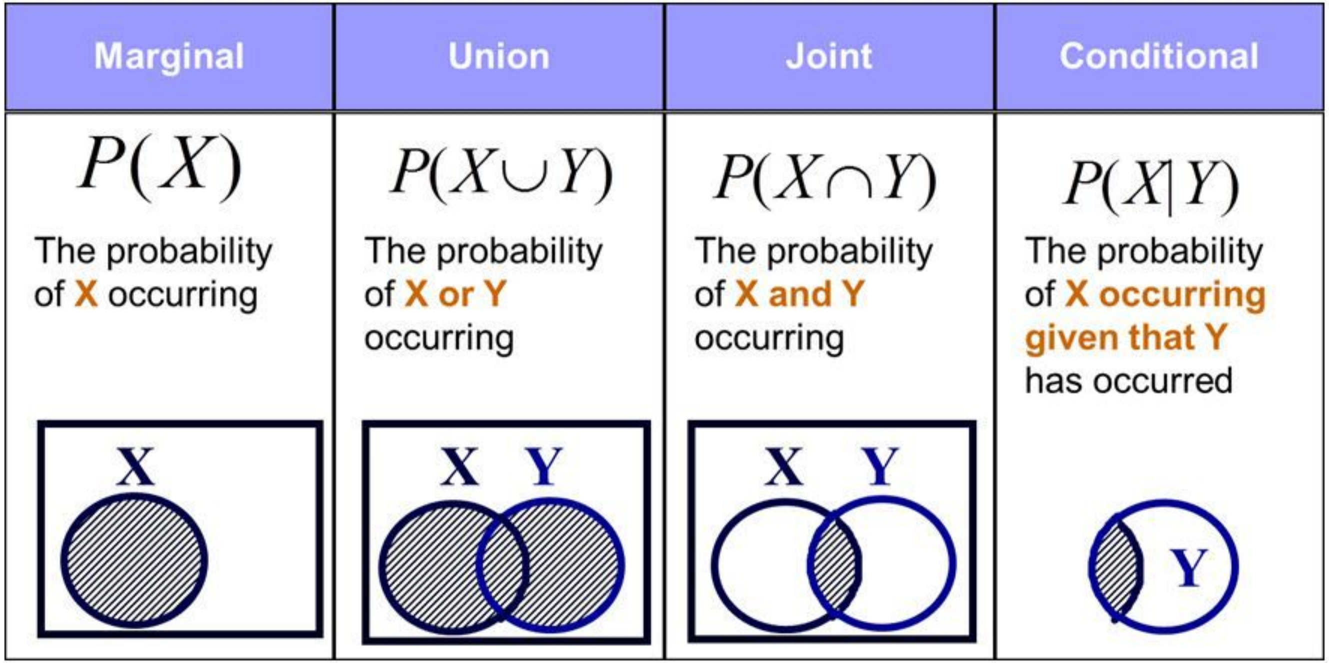
Outline

1. Set theory
2. Probability Formula
3. Marginal Probability
4. Joint Probability
5. Conditional Probability
6. Bayes Theorem
7. Bayesian Inference

Basic Set Operations



Types of Probabilities

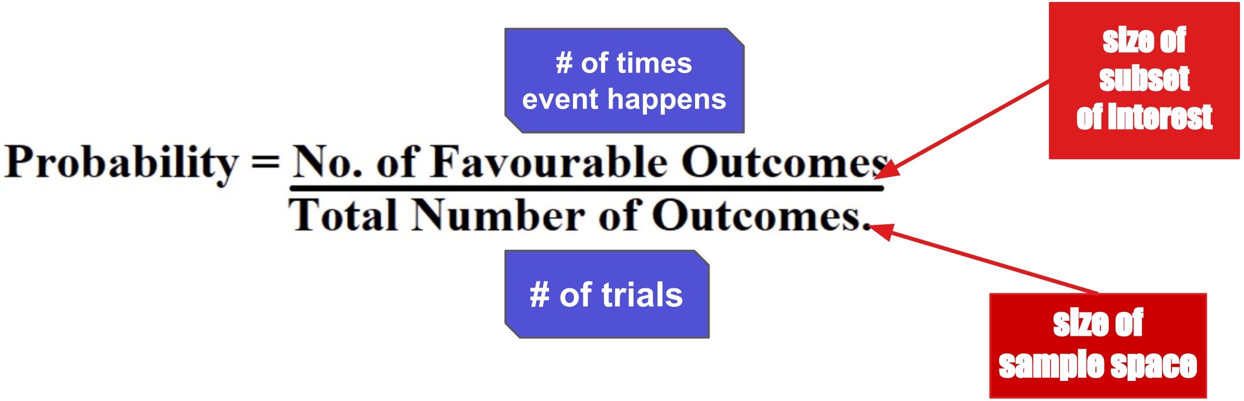


Marginal	Union	Joint	Conditional
$P(X)$	$P(X \cup Y)$	$P(X \cap Y)$	$P(X \mid Y)$
The probability of **X** occurring	The probability of **X or Y** occurring	The probability of **X and Y** occurring	The probability of **X occurring given that Y** has occurred

Probability of an event: a set of outcomes that is the subset of the sample space.

Probability Formula

- **a ratio that represents the likelihood of an event**
  - scientific way to measure uncertainty or randomness
    - → e.g. outcome- of tossing a coin; rolling a die; drawing a card;
    - → e.g. chance of- winning the lottery; defaulting on a loan; committing of fraud



$$\text{Probability} = \frac{\text{No. of Favourable Outcomes}}{\text{Total Number of Outcomes.}}$$

# of times event happens

# of trials

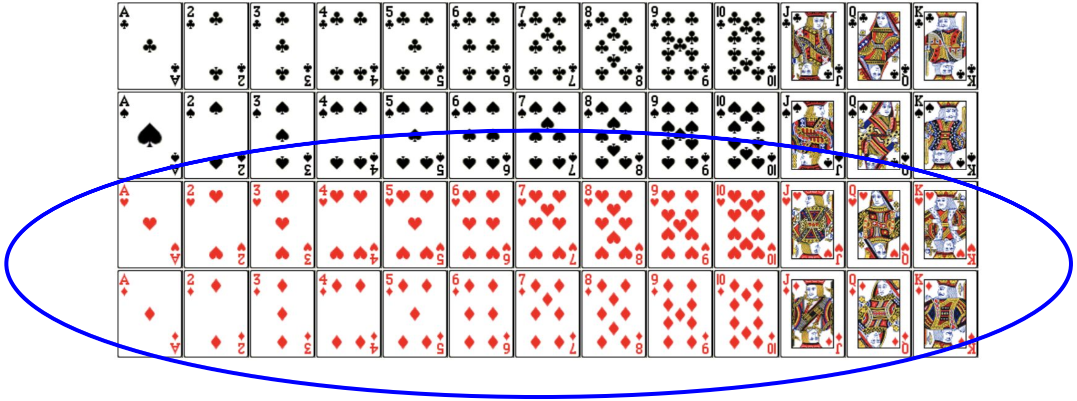
size of subset of interest

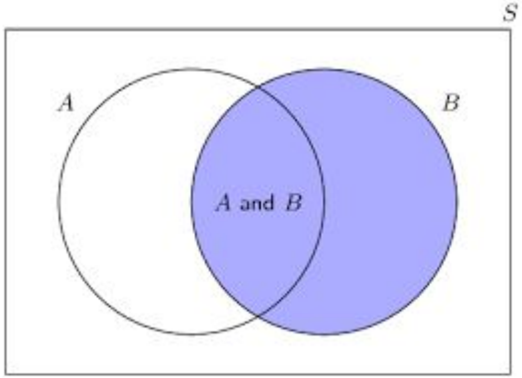
size of sample space

1. Marginal Probability

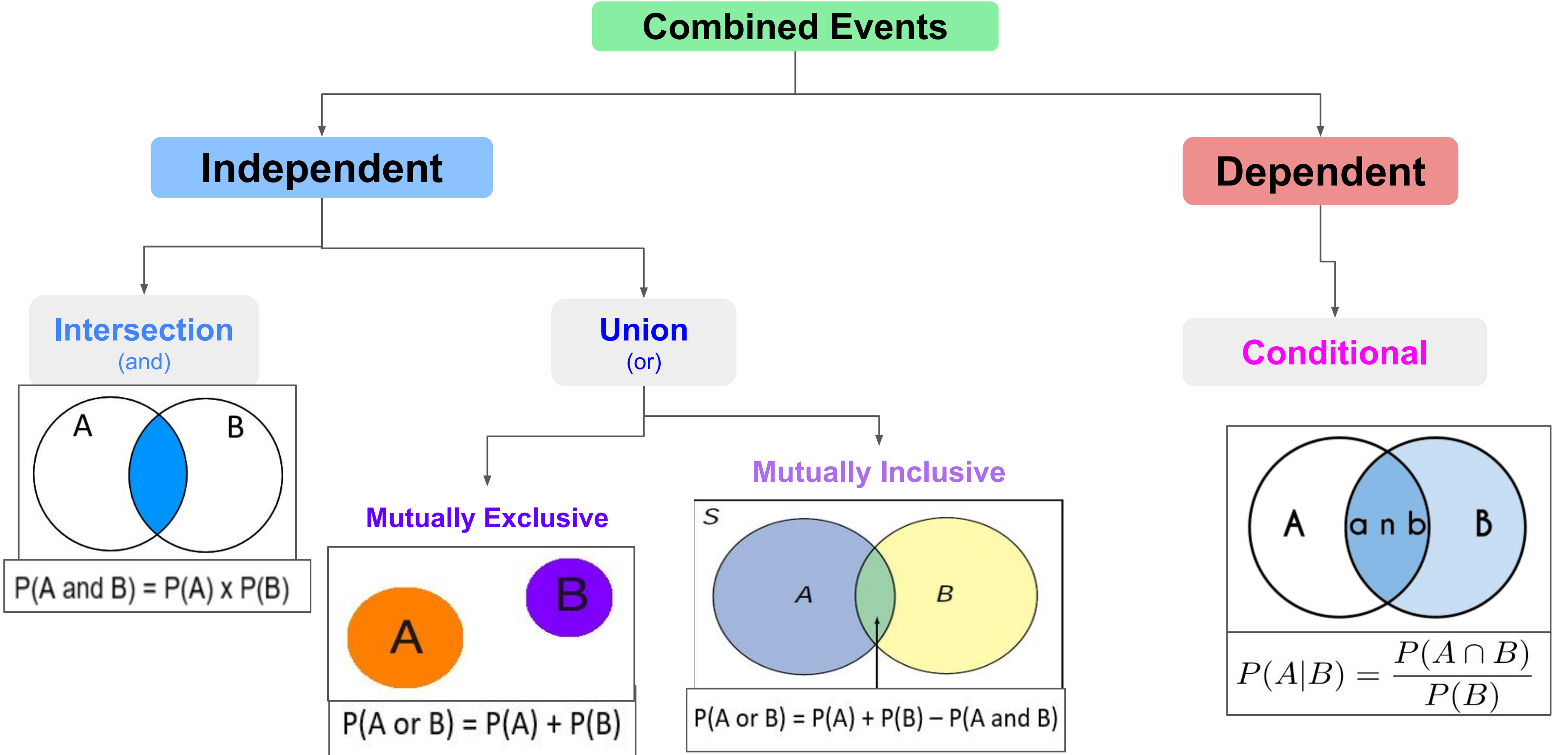
Marginal probability: the probability of an event occurring, P(A)

- thought of as an unconditional probability, not conditioned on another event.
    - e.g. the probability that a card drawn is Red, P(Red) = 26/52 = 0.5





Combined Events



Combined Events

Independent

Dependent

Intersection
(and)

P(A and B) = P(A) x P(B)

Union
(or)

Mutually Exclusive

P(A or B) = P(A) + P(B)

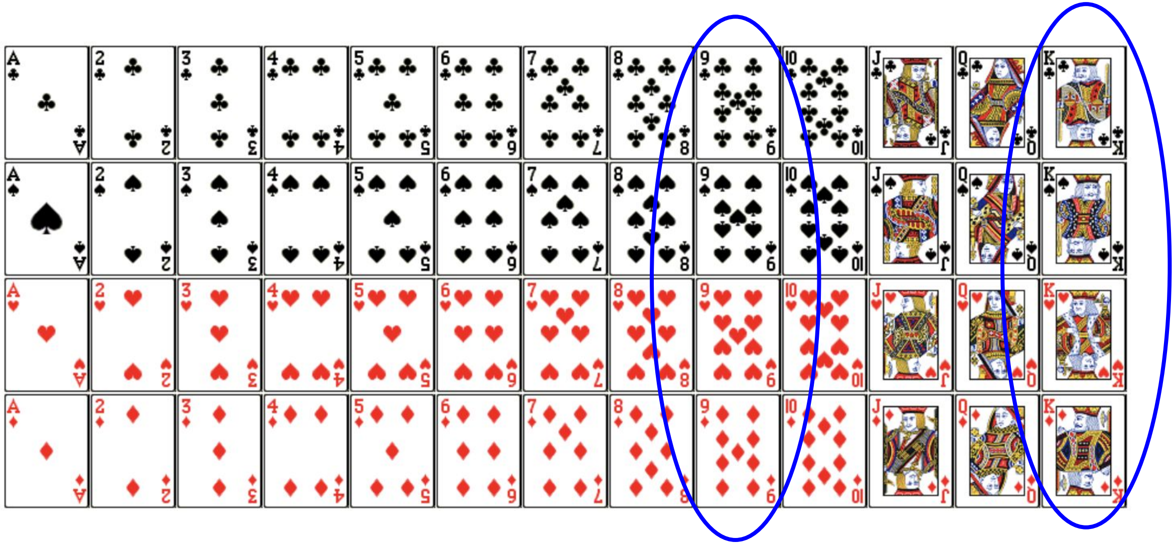
Mutually Inclusive

P(A or B) = P(A) + P(B) – P(A and B)

Conditional

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

2. Intersection

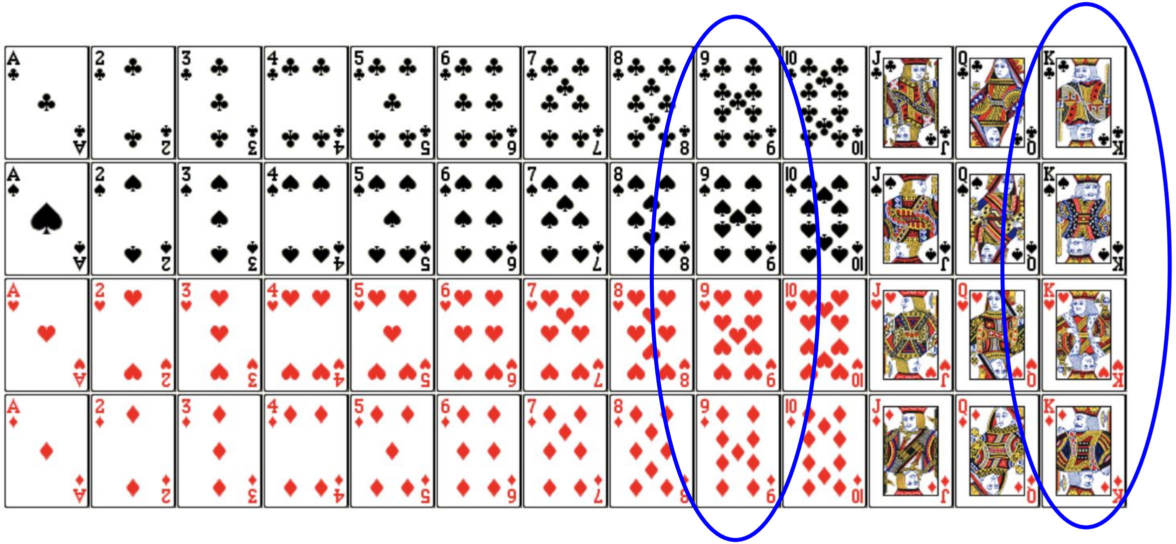


Two Events Occuring

- probability of drawing a **9** and a **King**
- P(**9** and **K**) = P(**9** ∩ **K**) = P(**9)** * P(**K**)

  = 4/52 * 4/52 = 16/ 2,704

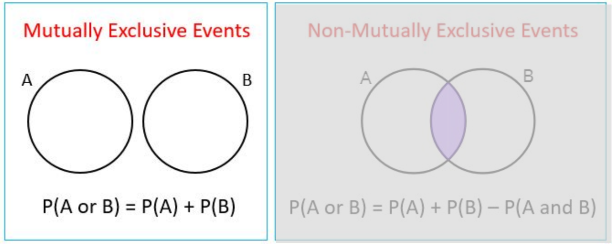
2.1 Union- Mutually Exclusive



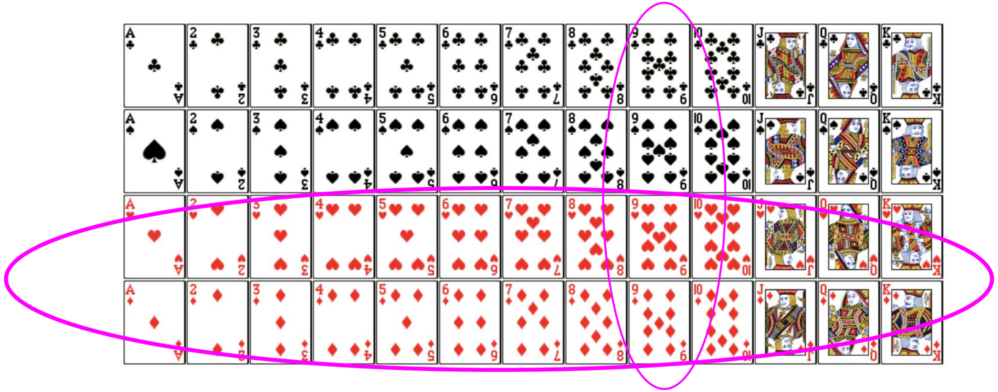
Mutually Exclusive Events

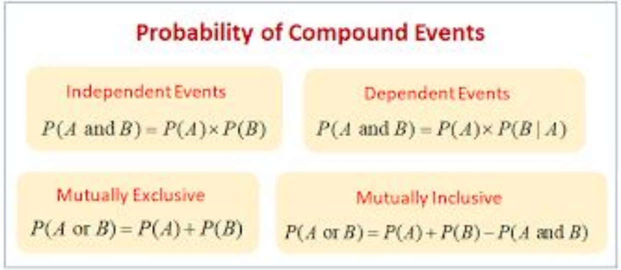
- Probability of drawing a **9** or **King**
- P(9 or K) = P(9) + P(K)

  = 4/52 + 4/52 = 8/ 52


Mutually Exclusive Events
A
B
P(A or B) = P(A) + P(B)
Non-Mutually Exclusive Events
A
B
P(A or B) = P(A) + P(B) – P(A and B)

2.2 Union - Mutually Inclusive





Probability of Compound Events

Independent Events

$P(A \text{ and } B) = P(A) \times P(B)$

Dependent Events

$P(A \text{ and } B) = P(A) \times P(B \mid A)$

Mutually Exclusive

$P(A \text{ or } B) = P(A) + P(B)$

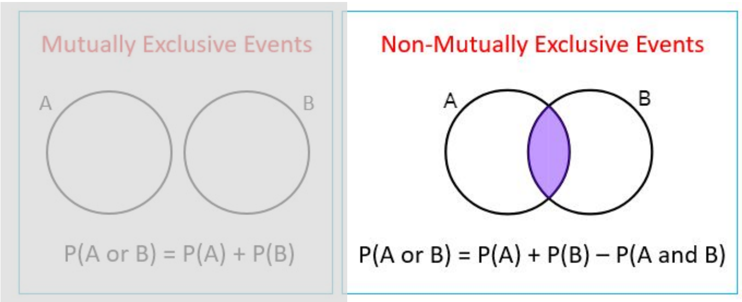
Mutually Inclusive

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Non-Mutually Exclusive Probability

- probability of drawing a **9** or a Red card
- P(Red or **9**)

  = P(Red) + P(**9**) - P(Red ∩ 9)
  = 26/52 + 4/52 - (1/2 * 4/52)
  = 28/52



Mutually Exclusive Events

P(A or B) = P(A) + P(B)

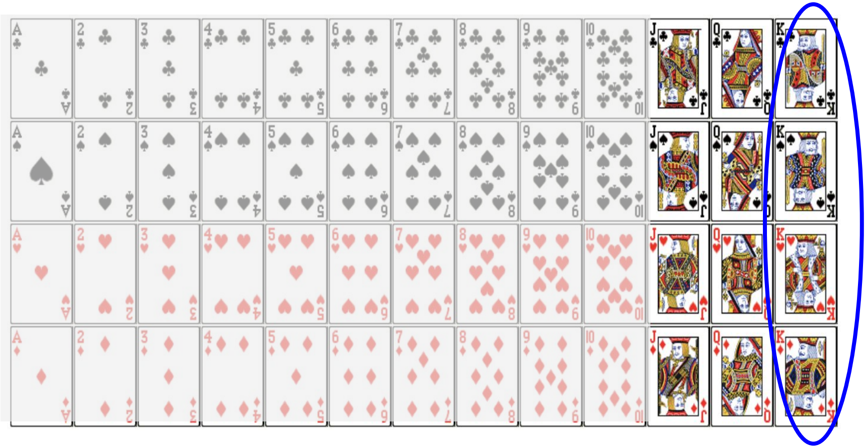
Non-Mutually Exclusive Events

P(A or B) = P(A) + P(B) – P(A and B)

3. Conditional Probability

the probability of event A, given that B occurred, P(A|B)

- derived from the probability multiplication rule, P(A and B) = P(A)*P(B|A)
    - given that you drew a **Face** card, what's the probability that it's a **King**



$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Dependent Probability

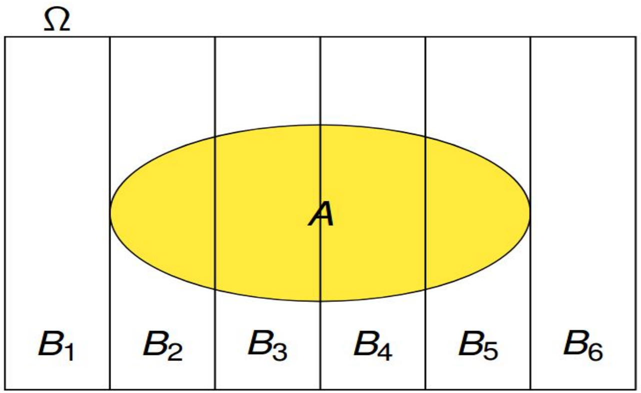
- P(**K | Face**) = P(**K∩Face**) / P(**Face**)

  = 4/52 / (12/52)

  = 4/12 = 1/3

Law of Total Probability

- **the total probability of an outcome realized via several distinct events**
  - a fundamental rule relating *marginal* probabilities to *conditional* probabilities
  - total probability of event A as a sum of conditional probabilities **B*i*** events
  - used to find the probability of an event, P(A), when P(A) cannot be calculated directly



$$P(A) = \sum_n P(A \cap B_n)$$

$$P(A) = \sum_i P(A|B_i)\, P(B_i)$$

$$P(A) = \sum P(A \cap B_x) = \sum P(A|B_x) * P(B_x)$$

$$P(A) = \sum P(A \cap B_x) = 0 + P(A \cap B_2) + P(A \cap B_3) + P(A \cap B_4) + P(A \cap B_5) + 0$$

Bayes Theorem

Bayes' Theorem

- **Studied logic & theology, 1719, Univ. of Edinburgh**
- **Motivated to rebut David Hume's argument**
  - against believing in miracles on the evidence of testimony
- **Findings on probability theory unpublished**
  - were passed to his friend Richard Price after his death (1761)

- **Solution to inverse probability problem**
  - published in the *Philosophical Transactions* of the RSL, 1973
  - popularized by Pierre-Simon Laplace, 1814
- **Use of a uniform prior distribution for a *binomial* parameter**
  - not general postulate,but now in all conditioned probabilities like SE, PML, RA
  - simultaneous localization & mapping, regularization, or information theory

Bayesian Inference

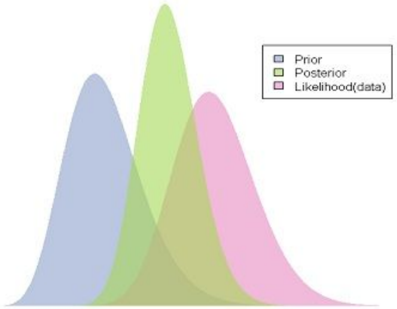
- **Spam Filtering**
  - attack spam by using words and phrases in the messages
  - determine the probability that the message is unwanted
- **Microsoft Operating System**
  - handwriting recognition; recommender systems; the question-answering
  - tracking business sales; preload the applications; predict traffic jam

- **Netflix $1-million contest**
  - online commerce relies on recommender filters
  - contest illustrates the prominent role of Bayesian
    - → Yehuda Koren, Christopher T. Volinsky, & Robert M. Bell
    - → winning algorithm had a Bayesian “perspective”
- **Google Searches**
  - several forms of Bayes to search through the billions of documents
  - Bayes’ filters through billions of pages, find **2** relevant ones in 0.31s

Bayesian Inference

- **update the probability of an event as per new evidence**
  - if the data support the hypothesis then the probability goes up
  - if it does not match, then probability goes down
- **adds ‘subjectivity’ to statistical modeling, ML, and AI**
  - knowledge discovery and predictive analytic of business, medicine, & technology
  - genetics, linguistics, image processing, brain imaging, epidemiology

Posterior distribution is proportional to data distribution * prior distribution

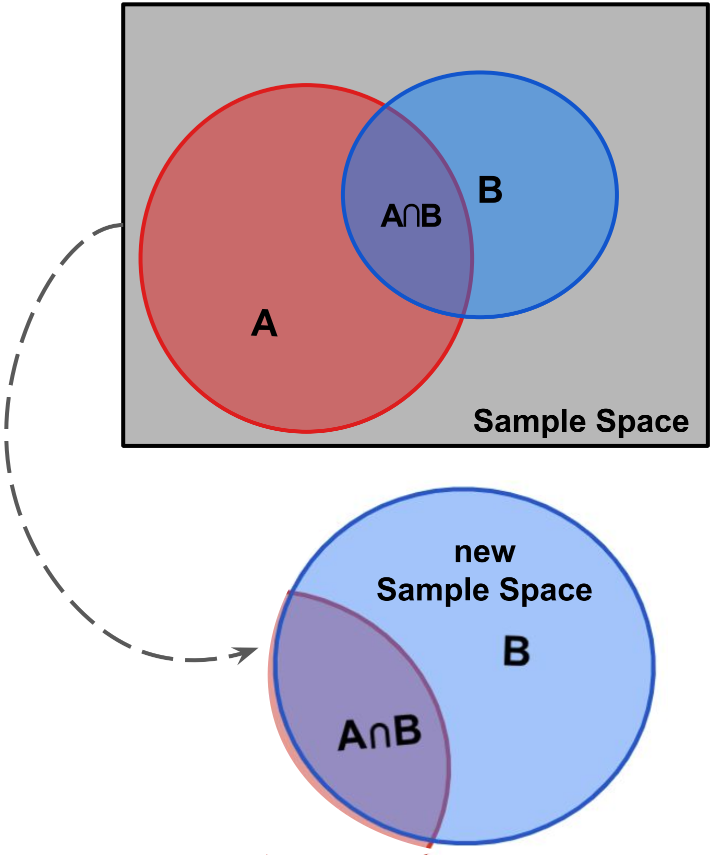


Bayes' Theorem

- **Determine P(A|B), *given* P(B|A)**
  - *updating probability with new info*
  - derived from rule of multiplication
    - → *if* P(A ∩ B) = P(A|B) P(B)
    - → *and* P(B ∩ A) = P(B|A) P(A)
    - → *then* P(B ∩ A) = P(A ∩ B)
  - P(A|B) P(B) = P(B|A) P(A)

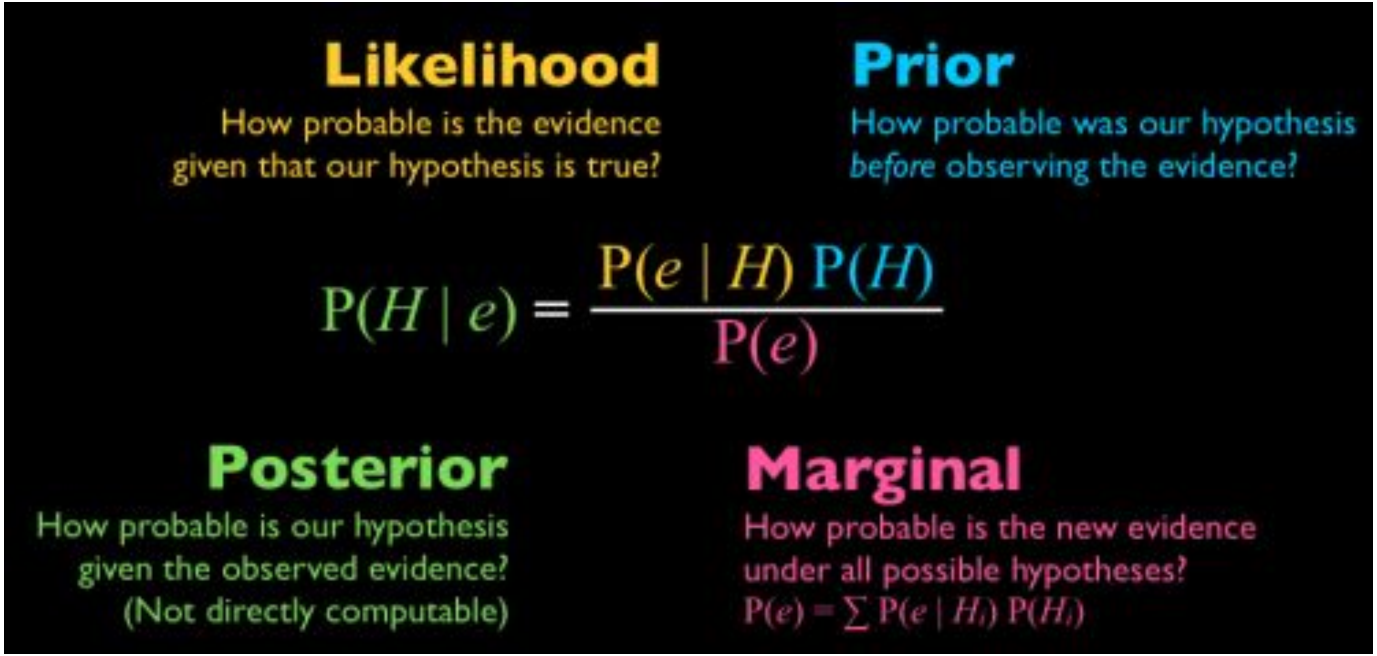
    → P(A|B) = P(B|A) P(A) / P(B)

$$P(B_J \mid A) = \frac{P(A \mid B_J)\,P(B_J)}{\sum_{i=1}^{n} P(A \mid B_i)\,P(B_i)}$$



Bayes’ Theorem

Posterior* = *likelihood* x *prior* / *new evidence



Likelihood

How probable is the evidence given that our hypothesis is true?

Prior

How probable was our hypothesis *before* observing the evidence?

$$P(H \mid e) = \frac{P(e \mid H)\, P(H)}{P(e)}$$

Posterior

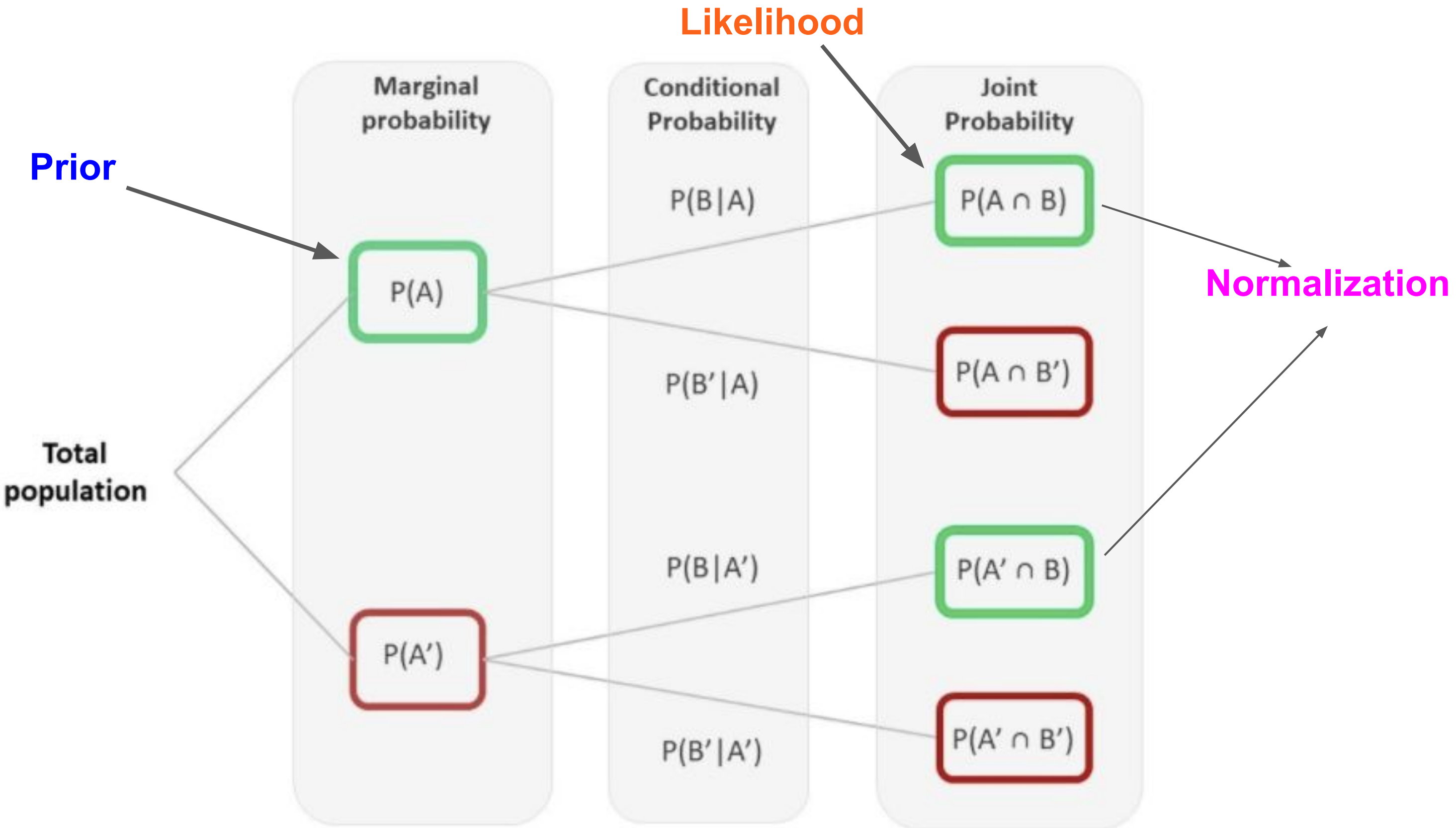
How probable is our hypothesis given the observed evidence?
(Not directly computable)

Marginal

How probable is the new evidence under all possible hypotheses?

$P(e) = \sum P(e \mid H_i)\, P(H_i)$

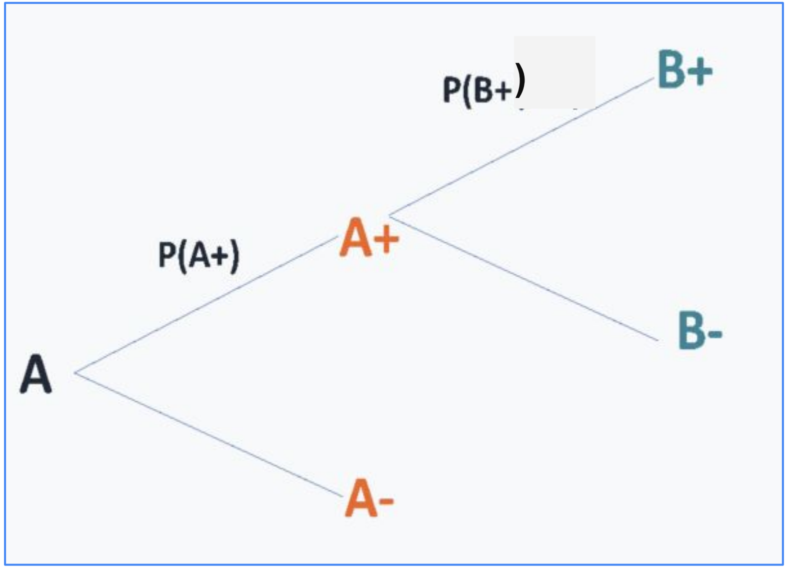
Bayesian Probability



Posterior* = *likelihood* x *prior* / *new evidence

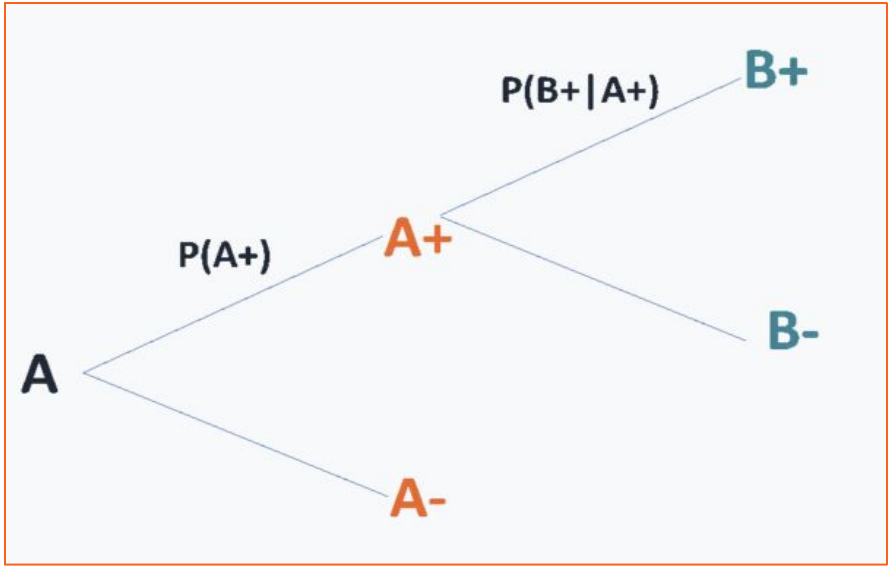
Conditional vs. Bayesian Theorem

Conditional Probability: the probability of event **B+** occurring ***in the presence*** of A+



How often does B+ happen when A+ happens?

Bayesian Probability: the conditional probability of event **A+**, ***given that*** event **B+** has already occurred



How often does A+ happen after B+ has happened?

Appendix

Applying Bayesian Inference

Application of Bayes- Rare Disease

A certain disease affects about 1 out of 10,000 people. There is a test to check whether the person has the disease. The test is quite accurate.

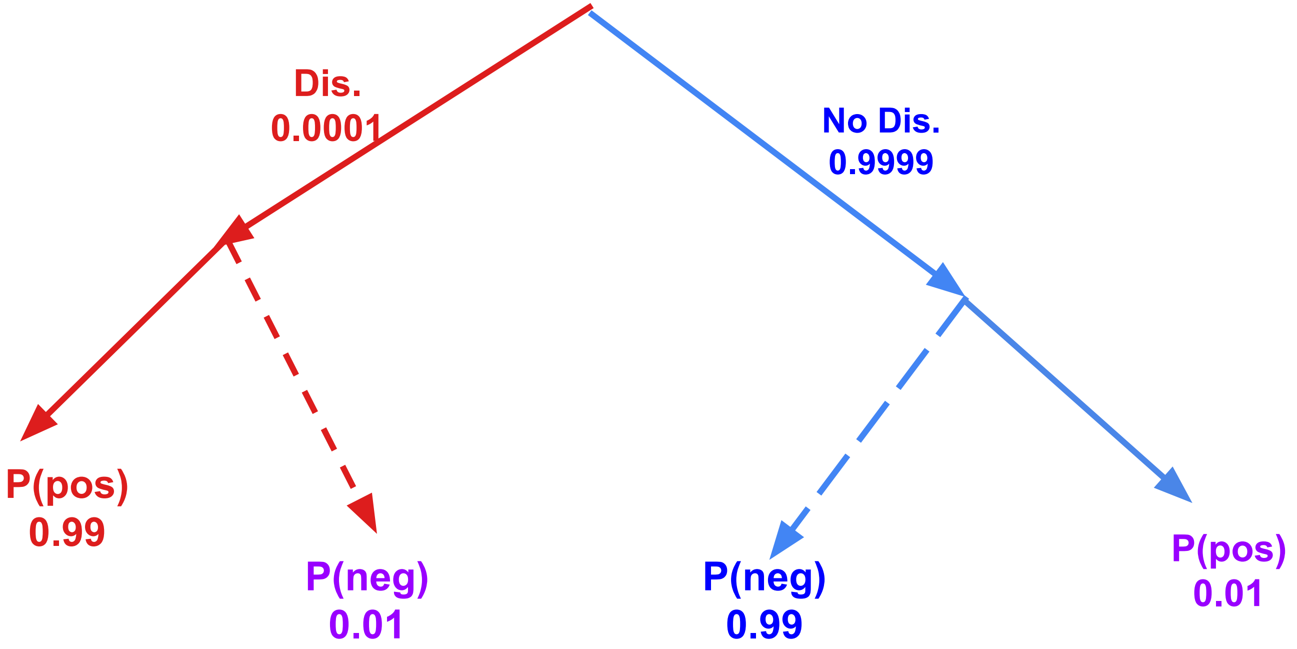
P(+) (suggesting the person has the disease), given that the person **does not have the disease**, is only 2 %;

P(-) (suggesting the person does not have the disease), given that the person **has the disease**, is only 1 %.

A random person gets tested for the disease and the result comes back positive.

What is the probability that the person has the disease P(D)?

Bayes Theorem Probability: Disease



$$P(sick|positive) = \frac{P(positive|sick)\ P(sick)}{P(sick)\ P(positive|sick) + P(not\ sick)\ P(positive|not\ sick)}$$

Application of Bayes- Rare Diseases

1. Let Dis be the event that the person has the disease
2. let Pos be the event that the test result is positive
3. We know: P(Dis|pos)
    - P(Dis) = .0001; P(NoDis) = 0.9999
    - P(pos|NoDis) = 0.01; P(neg|NoDis) = 0.99
    - P(neg|Dis) = 0.01; P(pos|Dis) = 0.99

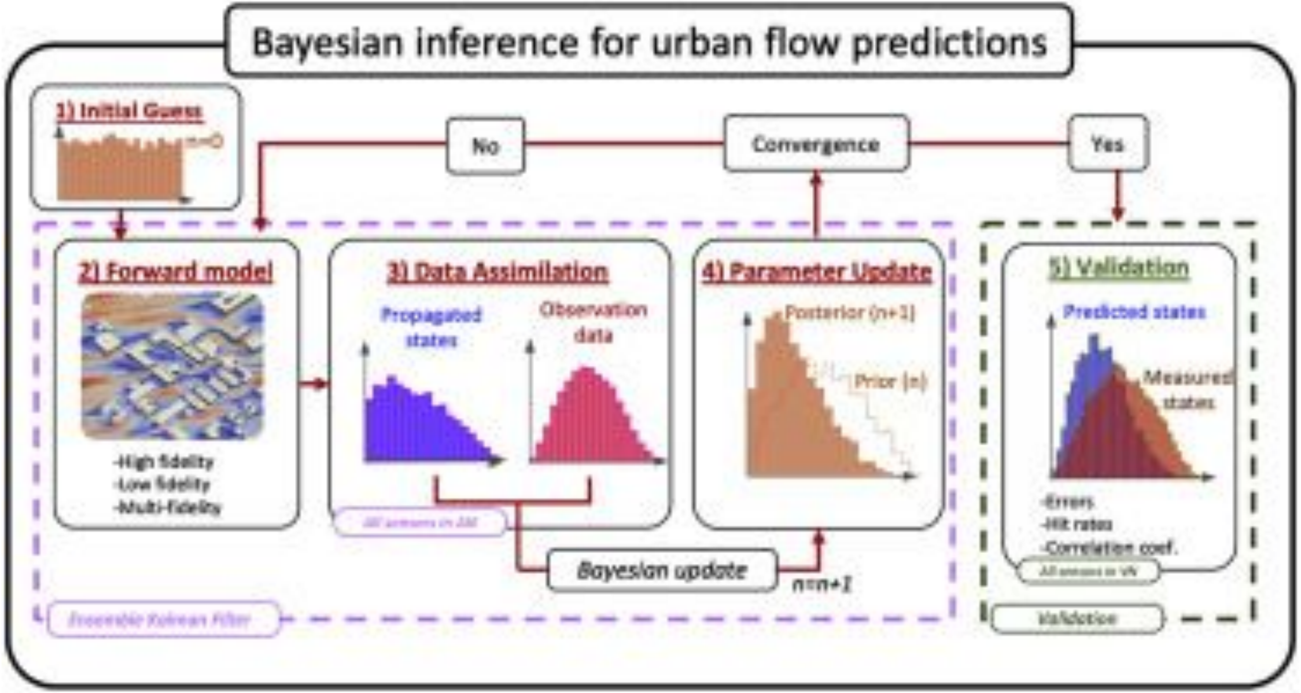
4. What we want to compute is P(D|T), we use Bayes' rule:
    - **P(Dis|pos)** = P(pos|Dis)**P(Dis)** / P(Pos|Dis)**P(Dis) + P(Pos|NoDis)P(NoDis)**

$$\frac{\mathbf{0.99 \times 0.0001}}{(\mathbf{0.99 \times 0.0001}) + (\mathbf{0.01 \times 0.9999})}$$

$$= \mathbf{0.0098}$$

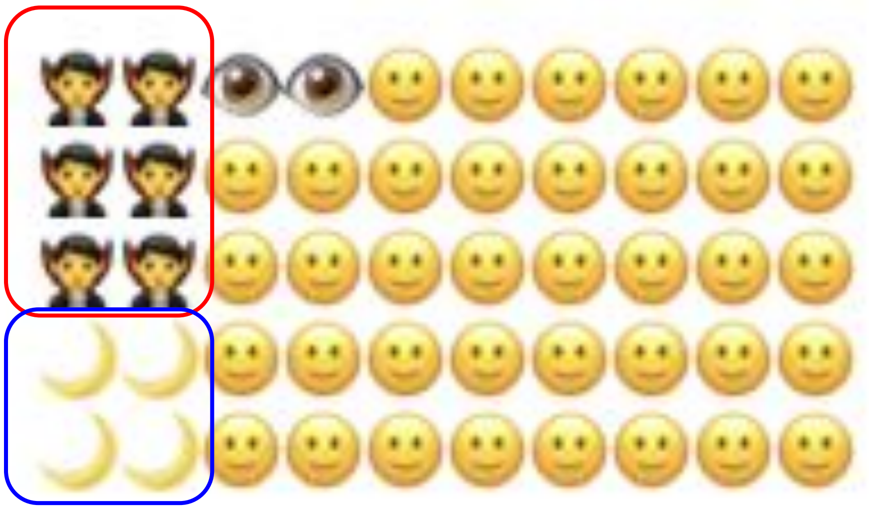
*This means that there is less than **0.98 %** chance that the person has the disease.*

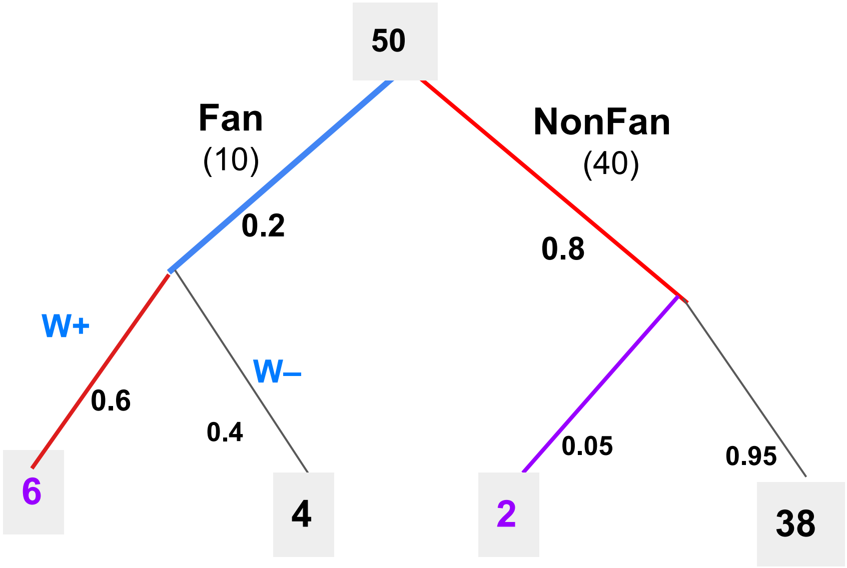
Bayesian Inference



Bayesian Inference

Sample of 50





6 Twilight Fans + 2 non Fans = 8 watched

P(F|W) = 6/8 = 0.75

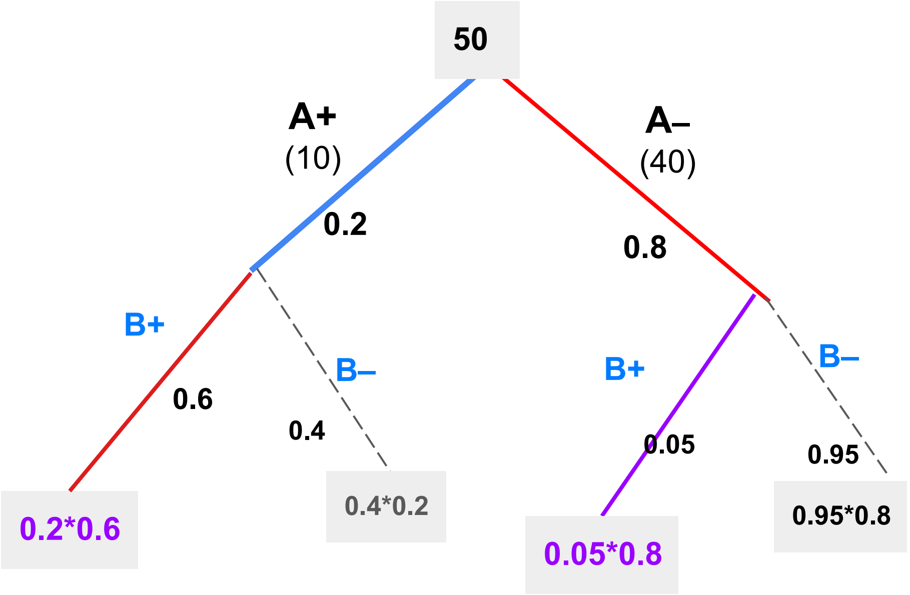
Bayesian Inference

P(fan) = P(A+) = 0.2
P(nonfan) = P(A-) = 0.2

P(watched) = P(B+)
P(watched) = P(B-)

P(watched given fan) = P(B|A) = 0.6
P(watched given nonfan) = P(B|A-) = 0.05

P(B) = P(B|A)*P(A) + P(B|A-)*P(A-) = 0.16



P(A|B) = P(B|A) * P(A) / P(B) = (0.12*0.2)/0.16 = 0.75

Bayesian Inference

	Yes	**No**	Total
Males	15	5	20
Females	17	3	20
Total	32	8	40

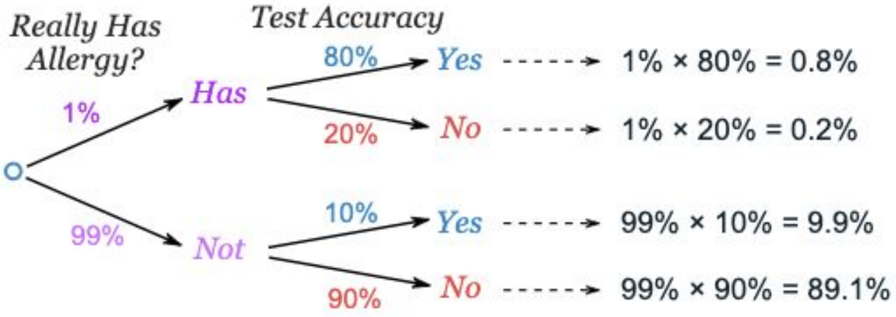
$$P\left(yes \mid male\right) = \frac{P\left(yes\ \ and\ \ male\right)}{P\left(male\right)} = \frac{15}{20}$$

Bayesian Probability

	1% have it	Test says "Yes"	Test says "No"
Have allergy	10	**8**	**2**
Don't have it	990	**99**	**891**
	1000	107	893

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

- P means "Probability of"
- | means "given that"
- A in this case is "actually has the allergy"
- B in this case is "test says Yes"



$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\text{not } A)P(B|\text{not } A)}$$

the probability of having the allergy given that the test is positive

$$= \frac{0.01\times0.8}{0.01\times0.8 + 0.99\times0.1}$$

$$= 0.0748\ldots$$